

# 编译原理课程实验 词法分析器

## 概述

你需要完成一个PL/0语言的编译器。这个实验分为若干部分。在这个部分中，你需要完成PL/0语言的词法分析器。

给定一个PL/0语言源程序，你需要将其从字符流转换为词语流。具体来说，你需要过滤源程序中的空白符(空格, tab, 换行等)，识别关键字、标识符、数字以及运算符。

PL/0的关键字表如下

1	CONST
2	VAR
3	PROCEDURE
4	BEGIN
5	END
6	ODD
7	IF
8	THEN
9	CALL
10	WHILE
11	DO
12	READ
13	WRITE

PL/0的标识符由上下文无关文法

1	<标识符> → <字母> {<字母>   <数字>}
2	<字母> → A   B   C ... X   Y   Z
3	<数字> → 0   1   2 ... 7   8   9

生成。

例如，ABC是合法的标识符，X25519是合法的标识符，但233QWQ不是合法的标识符

在PL/0语言中，标识符不会超过10字符长。如果超过了10个字符，你应该认为这是一个词法错误。

特别的，关键字不能是标识符。

PL/0的数字由上下文无关文法

1	<无符号整数> → <数字> {<数字>}
2	<数字> → 0   1   2 ... 7   8   9

生成。

PL/0仅支持无符号整数。因此，对于-123，你的词法应当将其分为两个词语-和123

PL/0的运算符表如下

```
1 | =
2 | :=
3 | +
4 | -
5 | *
6 | /
7 | #
8 | <
9 | <=
10 | >
11 | >=
```

你需要注意拼接两个字符组成的运算符。这些运算符中间不会存在空白符，例如，对于> =应当识别为两个单词>和=

PL/0的分隔符表如下

```
1 | ;
2 | ,
3 | .
4 | (
5 | )
```

请注意：PL/0语言不区分大小写。

通过将字符流转换为词语流，你将为下一个部分（语法分析器）做好准备。

我们建议你除存储需要自动评测时需要的信息外，存储一些额外信息。你将在语法分析器中使用诸如单词类型等信息。另请参阅旧版词法分析器中的GETSYM部分。如果你希望在未来的错误处理部分中提示错误信息，你可能还需要存储每一个词语在字符流中的位置。

## 评测要求

你可以选择自动评测，将这一部分的代码提交到SDU OJ上。通过自动测试后，这一部分的正确性分数将自动评为满分。

如果你不能或不愿进行自动评测，你可以要求人工验收。在这种情况下，这一部分的正确性分数将由助教给出。

如果你计划进行自动评测，你需要从标准输入读取输入的程序，并将分析得到的词语输出至标准输出，每行一个。

对于关键字，运算符，分隔符，直接输出其本身即可。

对于数字，你需要先输出一个单词NUMBER，之后是一个空格，之后是数字值。

对于标识符，你需要先输出一个单词IDENTIFIER，之后是一个空格，之后是标识符本身。

如果程序包含词法错误，你的程序应当仅输出一行"Lexical Error"，不含引号。

在本程序中，你只需要关注词法上的错误。对于词法正确而存在语法错误的程序，你仍然需要输出单词流。

输出时，除Lexical Error外，输出字母应当均为大写。

## 样例1

输入

```
1 VAR A,B;  
2 CONST C=0;  
3 BEGIN  
4     READ(B,A);  
5     A:=B+C;  
6     WRITE(A);  
7 END.
```

输出

```
1 VAR  
2 IDENTIFIER A  
3 ,  
4 IDENTIFIER B  
5 ;  
6 CONST  
7 IDENTIFIER C  
8 =  
9 NUMBER 0  
10 ;  
11 BEGIN  
12 READ  
13 (  
14 IDENTIFIER B  
15 ,  
16 IDENTIFIER A  
17 )  
18 ;  
19 IDENTIFIER A  
20 :=  
21 IDENTIFIER B  
22 +  
23 IDENTIFIER C  
24 ;  
25 WRITE  
26 (  
27 IDENTIFIER A  
28 )  
29 ;  
30 END  
31 .
```

## 样例2

输入

```
1 VAR  
2 THISISATOOLONGIDENTIFIER;
```

输出

```
1 | Lexical Error
```

### 样例3

输入

```
1 | VAR
2 | VARX = 233 = 0233;
```

输出

```
1 | VAR
2 | IDENTIFIER VARX
3 | =
4 | NUMBER 233
5 | =
6 | NUMBER 233
7 | ;
```

## 提交要求

---

你需要将代码提交到GitHub Classroom中。如果你选择自动评测，还需要提交代码以通过SDUOJ上的测试。

## 杂项

---

这一个实验部分建议完成时间约为两小时。